

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

**Automatización de procesos de captura de
información y gestión de datos digitalizados a partir de
motores de reconocimiento**

Franco Monreal Cortés

Profesor Guía: **Rodrigo Alfaro Arancibia**
Profesor Co-Referente: **Rodolfo Villarroel Acevedo**

Carrera: **Ingeniería ejecución informática**

Diciembre 2016

Índice

Resumen.....	ii
Abstract.....	ii
1 Alcances y propósitos del documento	1
1.1 Introducción	1
2 Descripción del proyecto	2
2.1 Descripción de la organización analizada.....	2
2.1.1 Personal de trabajo	2
2.1.2 Flujo actual sin sistema	3
2.1.3 Flujo con sistema.....	3
2.2 Problemas.....	4
2.3 Soluciones.....	4
3 Especificación de requerimientos.....	5
3.1 Objetivo General.....	5
3.2 Objetivos específicos	5
3.3 Requerimientos funcionales.....	5
3.4 Requerimientos no funcionales.....	6
3.5 Usuarios del sistema	6
3.5.1 Administrador.....	6
3.5.2 Director.....	7
3.5.3 Secretaria.....	7
4 Plan de trabajo	8
4.1 Metodología	8
4.2 Carta Gantt.....	9
4.3 Detalle de las actividades.....	11
5 Estudio de factibilidad	14
5.1 Factibilidad técnica	14
5.1.1 Motores OCR	14
5.1.2 Esquema básico de un algoritmo de OCR.....	15
5.1.3 Comparación con patrones	15
5.1.4 Servicios OCR.....	15
5.1.5 Combinación de escaneado y OCR.....	16
5.2 Factibilidad económica	16
5.3 Factibilidad operacional.....	17

5.4	Factibilidad legal.....	17
6	Análisis y gestión de riesgo.....	19
6.1	Riesgo de la tecnología a construir.....	19
6.2	Riesgo de las características del cliente.....	19
6.3	Riesgo asociado a la continuidad del software.....	19
7	Análisis de la solución.....	20
7.1	Arquitectura del sistema.....	20
7.2	Diagrama relacional.....	21
7.3	Diagrama de clases.....	22
7.4	Caso de uso general del sistema.....	22
7.5	Caso de uso administrador.....	23
7.6	Caso de uso director.....	24
7.7	Caso de uso secretaria.....	24
7.8	Caso de uso narrativo del sistema inicial.....	26
7.10	Caso de uso narrativo Módulo Secretaria.....	27
7.11	Caso de uso narrativo Módulo Administrador.....	28
7.13	Caso de uso narrativo Módulo Director.....	29
8	Diagramas de secuencia.....	30
8.1	Diagrama de secuencia Ingresar Usuario.....	30
8.2	Diagrama de secuencia Clasificar Documento.....	31
8.3	Diagrama de secuencia Buscar Documento.....	31
9	Prototipos gráficas.....	32
9.1	Interfaz de acceso al sistema.....	32
9.2	Interfaz módulo director.....	32
9.3	Interfaz secretaria.....	33
9.3.1	Clasificar documentos.....	33
9.3.2	Ver Perfil.....	34
9.4	Interfaz administrador.....	34
9.4.1	Administrar perfiles.....	35
9.4.2	Ver perfil de otro usuario.....	35
9.4.3	Agregar Usuario.....	36
9.4.4	Reportes de documentos.....	36
9.4.5	Estadísticas del sistema.....	37
10	Conclusión.....	38

11	Referencias	39
12	Anexos	40
12.1	Acta final de notas	40
12.2	Solicitud de pago de factura.....	41

Resumen

Este proyecto nace desde las dependencias del Instituto de Física de la Pontificia Universidad Católica de Valparaíso (IFIS) que necesita digitalizar archivos de manera rápida y automatizada. Por esto se propone crear un sistema el cual permita hacer la digitalización automatizada de un documento escaneado y procesado por algoritmos de reconocimiento de texto, organizando los datos de manera óptima para el fácil acceso a la información obtenida. El sistema debe contar con motores OCR que permitirán reconocer de buena manera el tipo de documento y sus atributos. Esto implica una importante mejora en la calidad de muchos servicios ahorrando al personal el quehacer de tediosas tareas. Este proyecto es específico y orientado a un departamento en particular, mas no está limitado a ser ampliado, muy por el contrario, pretende convertirse en un sistema que entregue soluciones informáticas a los problemas similares de todas las áreas administrativas de la PUCV o incluso de alguna otra institución.

Palabras claves: digitalización automatizada de documentos, motores OCR.

Abstract

The present Project starts in the facilities of the Physics institute of Pontificia Universidad Católica de Valparaíso (PUCV), which has the need of digitalizing its documents in a faster and automatized way. That is why, the proposal is to create a system, which would permit the automatized digitalization of a scanned documents, which would also be processed by text recognition algorithms. Additionally, it would organize data in an efficient and easier way to access to it. The system, should count with OCR motors that will allow the recognition of a document's type and its components. Therefore, it would imply an important improvement in the quality of several services through the avoidance of tedious everyday tasks. This project, is specific and its target audience would be a particular department, however; it may as well be extended. Conversely, this project pretends to be aggrandized to give informatics solutions to all direction areas in PUCV or even other institutions.

Keywords: documents automatic digitalization, OCR motors.

Lista de Figuras

Figura 2-1 Flujo actual del sistema.....	3
Figura 2-2 Flujo con un sistema informático	3
Figura 4-1 Modelo iterativo incremental.....	8
Figura 4-2 Carta Gantt etapa inicial	9
Figura 4-3 Carta Gantt etapa desarrollo	10
Figura 7-1 Arquitectura del sistema	20
Figura 7-2 Diagrama Relacional	21
Figura 7-3 Diagrama de clases	22
Figura 7-4 Diagrama de caso de uso general del sistema.....	23
Figura 7-5 Diagrama caso de uso módulo administrador.....	24
Figura 7-6 Diagrama caso de uso módulo director.....	24
Figura 7-7 Diagrama caso de uso módulo secretaria.....	25
Figura 8-1 Diagrama de secuencia iniciar sesión	30
Figura 8-2 Diagrama de secuencia clasificar/reportar documento	31
Figura 8-3 Diagrama de secuencia buscar documento	31
Figura 9-1 Acceso al sistema.....	32
Figura 9-2 Interfaz director.....	33
Figura 9-3 Interfaz clasificar documentos secretaria.....	33
Figura 9-4 Interfaz ver perfil	34
Figura 9-5 Interfaz administrar perfiles administrador.....	35
Figura 9-6 Interfaz ver perfil de secretaria	35
Figura 9-7 Interfaz agregar usuario	36
Figura 9-8 Interfaz de documentos reportados	36
Figura 9-9 Interfaz estadísticas del sistema.....	37
Figura 12-1 Acta final de notas digitalizada.....	40
Figura 12-2 Solicitud de comprobante de pago digitalizada	41

Lista de tabla

Tabla 3-1 Objetivos específicos	5
Tabla 3-2 Requerimientos funcionales	5
Tabla 3-3 Requerimientos no funcionales	6
Tabla 4-1 Etapa de planificación	11
Tabla 4-2 Etapa de análisis de requerimientos	11
Tabla 4-3 Etapa de análisis de sistema	12
Tabla 4-4 Etapa de diseño del sistema.....	12
Tabla 4-5 Etapa final de la primera fase.....	12
Tabla 4-6 Etapa de desarrollo de software	12
Tabla 4-7 Etapa de pruebas de software.....	13
Tabla 4-8 Etapa de análisis de resultados y conclusiones	13
Tabla 5-1 Descripciones de hardware	14
Tabla 5-2 Descripción de software	14
Tabla 5-3 Detalle de insumos	17
Tabla 7-1 Caso de uso narrativo del sistema inicial	26
Tabla 7-2 Caso de uso narrativo módulo secretaria	27
Tabla 7-3 Caso de uso narrativo módulo administrador	28
Tabla 7-4 Caso de uso módulo director.....	29

1 Alcances y propósitos del documento

1.1 Introducción

Podemos observar que la tecnología avanza de manera exponencial y con eso deberían avanzar de igual manera los procesos de automatización con el propósito de evitar cualquier trabajo humano que sea posible hacer por computadora. Existen muchas empresas, instituciones y asociaciones que aún gestionan todos sus procesos de forma manual, esto implica que se generen continuamente grandes cantidades de documentos con información relevante, teniendo que llenar bodegas de papeles los cuales después de un tiempo empiezan a perder legibilidad llegando hasta el punto de deteriorarse por completo la información registrada.

El Instituto de Física (IFIS) de la Pontificia Universidad Católica de Valparaíso es una de las instituciones que viene lidiando con estos problemas hace ya un tiempo, esta institución cuenta con una estantería llena de archivadores con documentos que van desde el año 2000 hasta el presente, el gran problema es cuando algún trabajador necesita rescatar datos de algún documento, resulta imposible realizar una búsqueda rápida teniendo que realizar la tediosa tarea de buscar archivador por archivador y luego documento por documento hasta encontrar el dato solicitado. Para solucionar este problema la institución adquirió una impresora multifuncional la cual cuenta con un sofisticado escáner que permite tener los archivos de manera digital, aun así, podemos notar un gran trabajo humano en lo que es la digitalización manual de los datos, teniendo que hacer un enorme esfuerzo y un costo trabajo escaneando papel por papel y traspasando los datos de forma lenta e ineficiente. Debemos tener en cuenta que hoy en día existen herramientas de reconocimiento que simplificarían estas tareas permitiendo automatizar la captura de información evitando entradas por teclado de parte del usuario. Es por esto que el IFIS ha expresado la necesidad de un sistema el cual cumpla la función de automatizar los procesos de captura de información y gestionar los datos digitalizados utilizando técnicas informáticas de reconocimiento de textos.

El proyecto tendrá resultados significativos para una institución donde el trabajo más tedioso es el de la organización de documentos, mejorando significativamente los tiempos de trabajo. Además los usuarios podrán organizarse de mejor manera para que cada uno digitalice los documentos que necesite, aportando en mejorar el ambiente laboral y separando el quehacer de cada usuario en tareas específicas.

2 Descripción del proyecto

En el siguiente proyecto de reconocimiento de texto y automatización de proceso de captura de información se propone generar un sistema el cual permita digitalizar documentos físicos a través de un escáner de alto rendimiento. Al momento de escanear el documento se obtiene una imagen digital en formato JPG a la cual se le hará extracción de la información para identificar el tipo de documentos y sus atributos para posteriormente traspasarlos a las bases de datos y almacenarlos en los módulos correspondientes.

Dentro del IFIS los documentos almacenados se dividen en dos grandes grupos por una parte se encuentra el módulo de dirección que gestiona todo los archivos relacionados con antecedentes curriculares, facturas, seminarios, etc. Donde el usuario encargado de administrar este módulo es el personal que trabaja en el área de dirección. Por otro lado tenemos el módulo de docencia, encargado de registrar todo documento relacionado con los alumnos como actas de notas, certificados de calificación, encuesta docente, etc. Donde el usuario que gestiona este módulo es la asistente de docencia.

2.1 Descripción de la organización analizada

El instituto de física de la PUCV es la academia que tiene como misión el cultivo, el desarrollo y la difusión de la ciencia física a través de sus alumnos y sus académicos para, de esta forma, garantizar el adecuado conocimiento de esta disciplina. De esta forma el instituto contribuye a la generación de profesionales y graduados en diversas áreas del saber, otorgándoles una sólida base científica que les permite un adecuado desempeño bajo los más variados ámbitos y condiciones además de una formación con vocación de servicio a la sociedad.

2.1.1 Personal de trabajo

Dentro de las personas que trabajan en el instituto de física podemos encontrar académicos y administrativos. Entre los académicos se encuentran profesores de vasta categoría con estudios de doctorados y postgrados que se encargan de entregar la educación necesaria a los estudiantes. Por otra parte están los administrativos del instituto –a quienes está orientado este proyecto-, principalmente está dirigido a la secretaria de dirección y a la secretaria de docencia. También está dirigido a otros importantes entes que trabajan allí; el director del instituto y el administrador de redes quien será el encargado gestionar el sistema informático.

2.1.2 Flujo actual sin sistema

El flujo actual sin un sistema informático se reduce en la compleja tarea que tienen las secretarías de docencia y administración en la digitalización manual de los documentos, luego de obtener los archivos de manera digital las secretarías deben ir al computador para traspasar los archivos a los módulos correspondientes de forma manual. Para explicar de mejor manera el flujo sin sistema se muestra la representación en el siguiente diagrama:

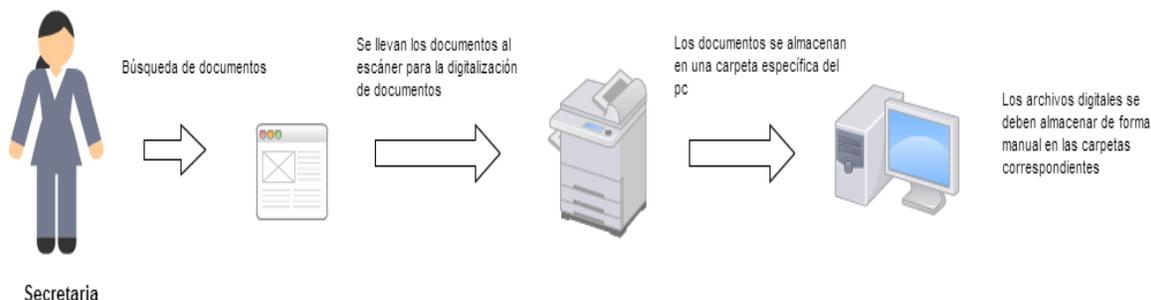


Figura 2-1 Flujo actual del sistema

2.1.3 Flujo con sistema

El flujo con un sistema informático que cuenta con el servicio de motores OCR queda representado en el siguiente diagrama donde podemos ver la mejora del trabajo de las secretarías. De igual forma se debe ocupar el escáner pero cuando llega el momento de clasificar los archivos ésta tarea se hace de forma automatizada, dando un solo clic para que los archivos a través de técnicas de reconocimiento de texto logren almacenarse en los módulos correspondientes. Para la seguridad de la información almacenada se usan bases de datos de alto nivel que permitirán el fácil y rápido acceso a la información. Además podemos notar que teniendo un sistema informático podrán tener acceso a los archivos otro tipo de usuarios como son el director del instituto y el administrador de redes.

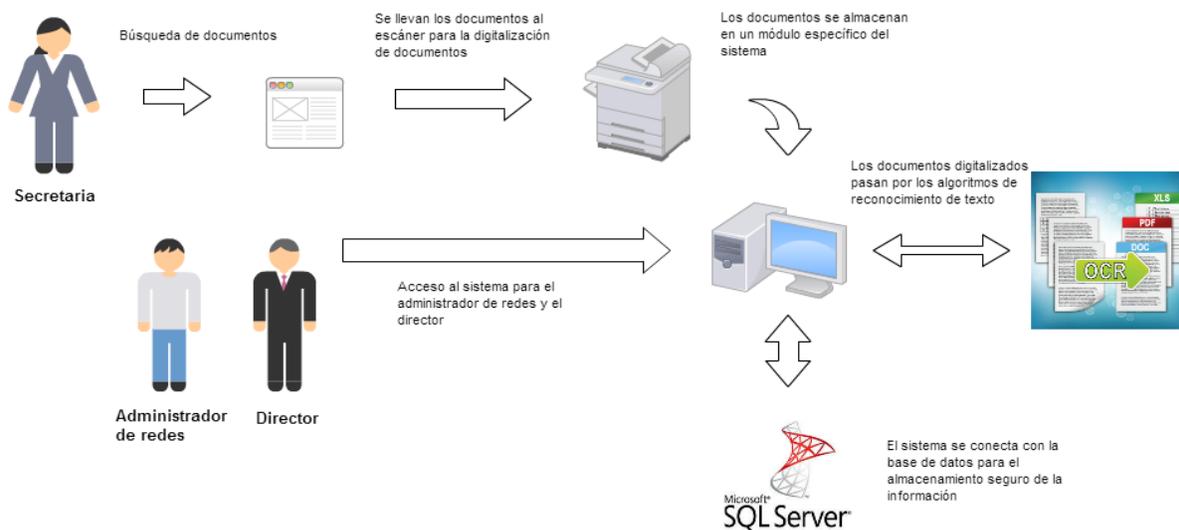


Figura 2-2 Flujo con un sistema informático

2.2 Problemas

- El instituto de física almacena una gran cantidad de documentos acumulándolos en archivadores los cuales se usan constantemente, dificultando la búsqueda rápida de algún dato específico, además existe la preocupación por el deterioro o pérdida de la información de los documentos.
- La digitalización de los documentos se hace a través de un escáner de alto rendimiento, el cual rescata la imagen y la guarda en una carpeta específica lo que implica que al momento de almacenar los documentos en los módulos correspondientes sea de manera lenta, ya que deben ser ordenados de forma manual.
- Se produce una mala organización entre los roles de las secretarías, pues ambos módulos de los documentos (docencia y dirección) son digitalizados solo por la asistente de dirección y no por la secretaria correspondiente a cada departamento.

2.3 Soluciones

- Para solucionar el problema de las estanterías sobre-cargadas de archivadores llenos de documentos se propone crear un sistema que automatice la digitalización de documentos a través de reconocimientos de patrones para la buena organización de los archivos y el rápido acceso a la información, esto implicará realizar un respaldo ordenado de dichos documentos de forma digital para evitar las pérdidas de datos o documentos físicos ya sea por daños, desgaste de papel, robo, etc.
- Respecto al problema de la digitalización manual de los documentos se propone que el sistema se encargue de los documentos escaneados y que a través del reconocimiento de texto haga de manera automatizada el almacenamiento de dicho documento en el módulo correspondiente.
- Se propone que el sistema esté protegido por un menú de acceso donde cada usuario podrá ingresar para poder digitalizar sus propios documentos en el sistema.

3 Especificación de requerimientos

La finalidad de este sistema es lograr satisfacer las necesidades de los usuarios los cuales plantearon en varias reuniones sus requerimientos. En el sistema se pueden destacar las siguientes funciones generales de este proyecto.

3.1 Objetivo General

Desarrollo de un prototipo que utiliza algoritmos de reconocimientos de textos para procesar datos de documentos físicos y gestionar la información digitalizada.

3.2 Objetivos específicos

Tabla 3-1 Objetivos específicos

OBJ 1	Automatizar procesos de captura de información a través del reconocimiento de texto utilizando motores OCR.
OBJ 2	Desarrollar un sistema de información que logre interactuar con los algoritmos de reconocimiento de texto.
OBJ 3	Lograr una tasa de aciertos en el reconocimiento de texto de 70% mínimo.
OBJ 4	Gestionar el sistema de reconocimientos para dos tipos de documentos distintos en primera instancia.

3.3 Requerimientos funcionales

Tabla 3-2 Requerimientos funcionales

RF1	Crear en el sistema general dos módulos, cada uno con sus sub-módulos los cuales se describen a continuación: <ul style="list-style-type: none">• Módulo dirección: Este módulo es responsable de almacenar todo tipo de documento relacionado al departamento de dirección, incluye documentos relacionados a solicitudes de pago, antecedentes curriculares, facturas, seminarios y congresos, entre otros.• Módulo de docencia: Este módulo está encargado de recopilar todo tipo de documento relacionado con los alumnos del instituto de física, estos documentos pueden ser actas de notas, certificados de calificación, notas de examen de grado, bases de concursos, etc.
RF2	El sistema debe contar con un método de autenticación de usuario que logre identificar a través del rut y contraseña, reconociendo el módulo de acceso a ingresar. Cuando el usuario se identifica el sistema reconoce el tipo de usuario y le da los privilegios correspondientes.
RF3	Crear un perfil de secretaria para acceder a los archivos del área de docencia y de dirección con funcionalidades que serán: buscar un documento específico y al encontrarlo poder verlo, eliminarlo, guardarlo o imprimirlo.
RF4	El perfil de secretaria tiene la opción de clasificar un documento analizado por el motor OCR, donde al clasificar un documento se aceptan los datos entregados por el motor de reconocimiento.

RF5	El perfil de secretaria en el caso de encontrar un archivo mal reconocido, puede reportarlo. Esta acción envía el documento digitalizado al módulo de reportes, en donde el administrador del sistema puede analizar los problemas que tiene y clasificarlo de buena manera.
RF5	Crear un perfil para el director del instituto de física. Este usuario tendrá acceso al sistema solo con opciones de lectura, búsqueda, imprimir y guardar documentos en sistema.
RF6	Crear un perfil de administrador del sistema, este usuario lo manejará el encargado del departamento de informática del instituto de física, cuenta con las funcionalidades para administrar los perfiles y los permisos del sistema: creación de usuarios, modificación de usuarios, entrega de permisos a un determinado usuario.
RF7	Al administrador le llegarán los reportes hechos en el perfil de secretaria de los archivos mal escaneados, donde él podrá analizar las causas y motivos por la que no se hizo correctamente la digitalización de la información, para modificarla antes de ingresarla a las bases de dato.
RF8	El administrador del sistema tendrá acceso a las estadísticas relacionadas con la cantidad de archivos escaneados por tipo de documento. Para esta opción se generan representativos gráficos con la información del sistema. Además tendrá acceso a analizar al usuario que más documentos ha escaneado.

3.4 Requerimientos no funcionales

Tabla 3-3 Requerimientos no funcionales

RNF 1	Se utilizará el lenguaje de programación Visual Basic .NET para la capa de presentación con una interfaz sencilla para el uso de usuarios principiantes.
RNF 2	El sistema es una aplicación de escritorio para sistemas operativos Windows con rendimientos óptimos en la realización de sus funciones.
RNF 3	El motor de reconocimiento a utilizar es Tesseract perteneciente a Google, el cual cuenta con una alta tasa de aciertos en el reconocimiento.
RNF 4	El sistema utilizara base de datos SQL server. Con base de datos locales que estarán siempre disponibles.

3.5 Usuarios del sistema

El sistema posee tres actores que se relacionan en forma directa con él. La relación que tienen con el sistema, las funcionalidades, permisos, formas de acceso y registro se conocerán a continuación:

3.5.1 Administrador

El administrador del sistema es el usuario encargado de gestionar toda la información interna del programa, la principal función del administrador es agregar y modificar o eliminar perfiles. Además tendrá opciones de revisar estadísticas del sistema respecto al uso de algunas funcionalidades sobre los documentos escaneados y usuarios, también cuenta con un módulo de reportes cuando existan fallas en la digitalización de documentos.

3.5.2 Director

El director es el usuario con menos uso del sistema, éste solo puede realizar visualizaciones de los documentos y hacer simples operaciones de búsqueda, impresión y almacenar los documentos.

3.5.3 Secretaria

La secretaria es el usuario más activo del sistema, es el que realiza la principal función del sistema que es la digitalización de los documentos. También es la que más ocupará la información escaneada trabajando con los datos obtenidos. En caso de que se reconozca de mala forma un documento tiene la opción de reportarlo. Además puede ver su perfil con opciones de modificar sus datos o contraseña.

4 Plan de trabajo

4.1 Metodología

Es necesario reconocer que el software es un sistema complejo que continuamente está evolucionando a medida que el tiempo pasa, por eso, se hace imposible finalizar un producto por completo, por lo que se debe introducir una versión limitada del programa para cumplir la presión competitiva del mercado. Existen distintos modelos evolutivos de proceso de software, por lo que se tendrá que escoger el que más acomode a nuestro sistema. Los modelos más utilizados son:

- Modelo incremental: Este modelo combina elementos del modelo lineal secuencial con la filosofía interactiva de construcción de prototipos.
- Modelo espiral: Es el proceso evolutivo que conjuga la naturaleza iterativa de la construcción de prototipos.
- Modelo cascada: Este proceso se caracteriza por que de forma estricta, solo cuando se finaliza una fase puede comenzar la otra.

La metodología a utilizar en el proceso evolutivo del software es el modelo iterativo incremental el cual se caracteriza por ser creado en respuesta a las debilidades del modelo tradicional de cascada. Básicamente este modelo es un conjunto de tareas agrupadas en pequeñas etapas repetitivas, lo que nos va a permitir ir avanzando en paralelo de manera continua en distintos ámbitos del proyecto. La siguiente imagen sirve para ejemplificar el modelo iterativo incremental.

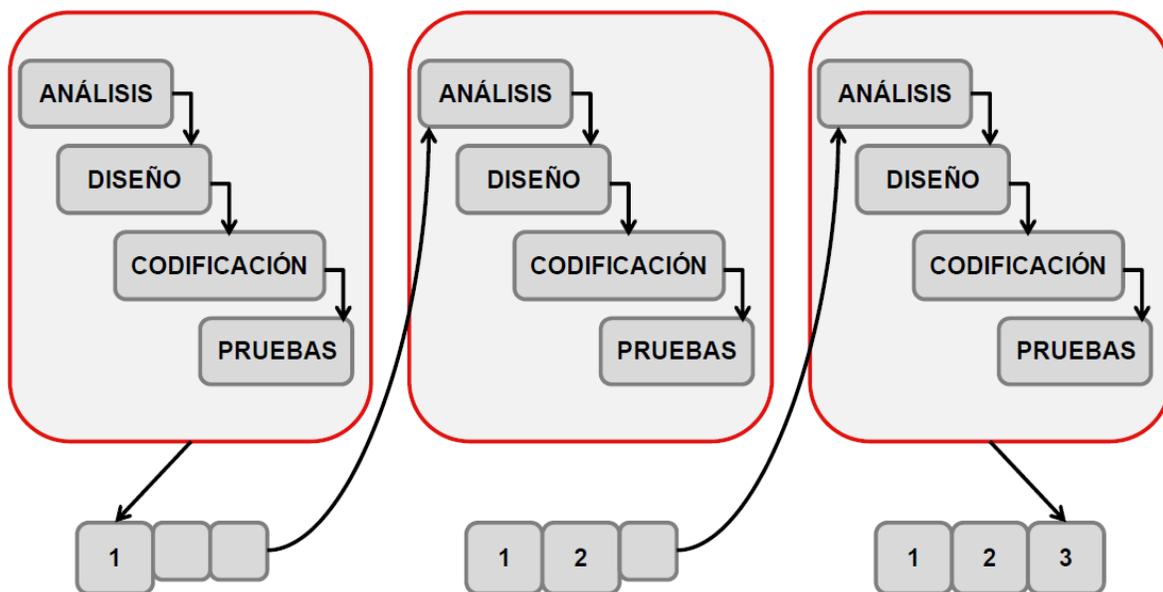


Figura 4-1 Modelo iterativo incremental

4.2 Carta Gantt

El siguiente diagrama de Gantt nos muestra de manera gráfica la dedicación para las distintas tareas o actividades a realizarse en las fechas y plazo establecidos. En primer lugar se ejecuta la etapa de planificación que consta de reuniones con el profesor guía y entrevista con los clientes que entregaron las bases para definir el proyecto. Luego en la fase de desarrollo se crean las funciones principales del sistema y finalmente en la fase de pruebas para analizar los resultados obtenidos del sistema.

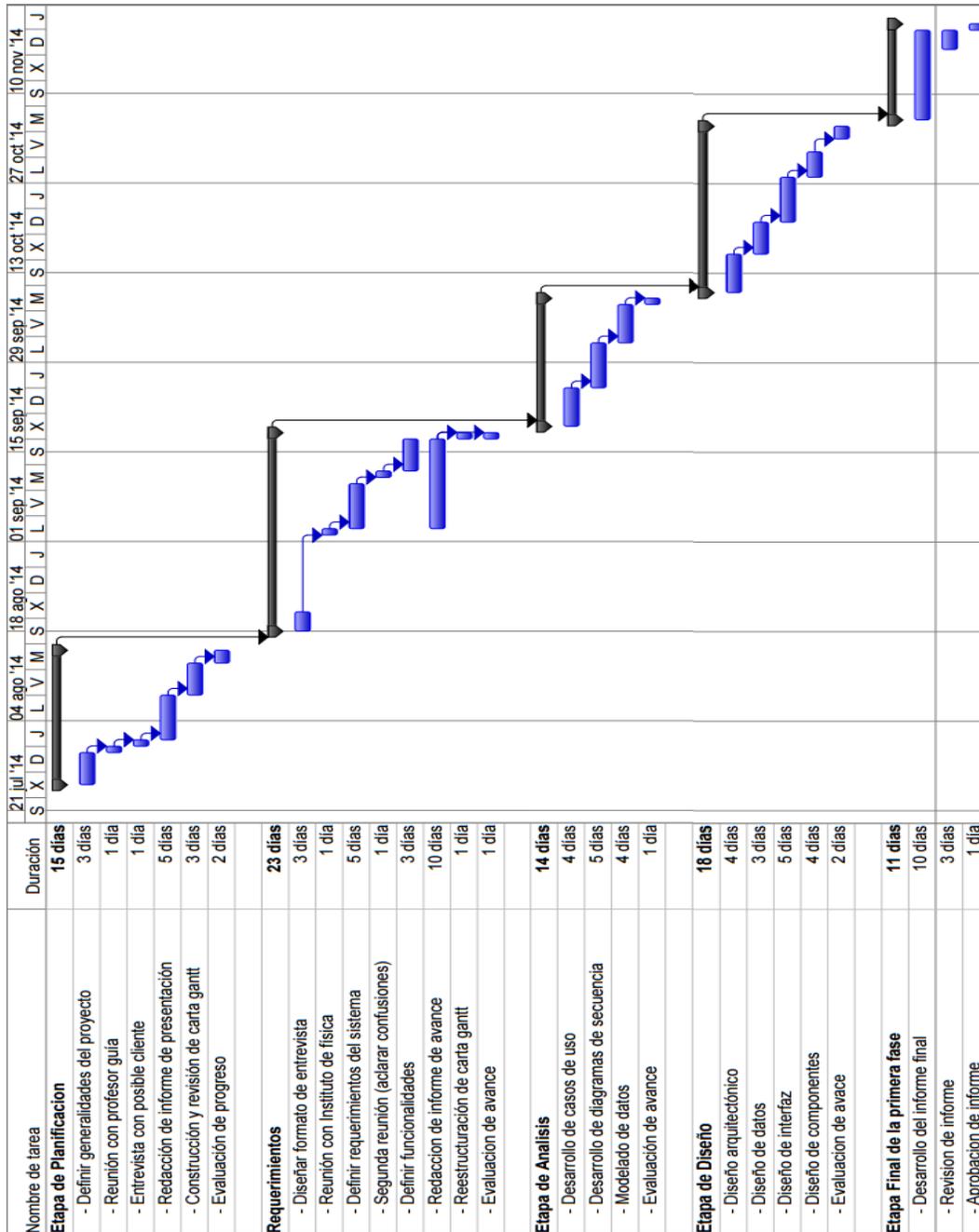


Figura 4-2 Carta Gantt etapa inicial

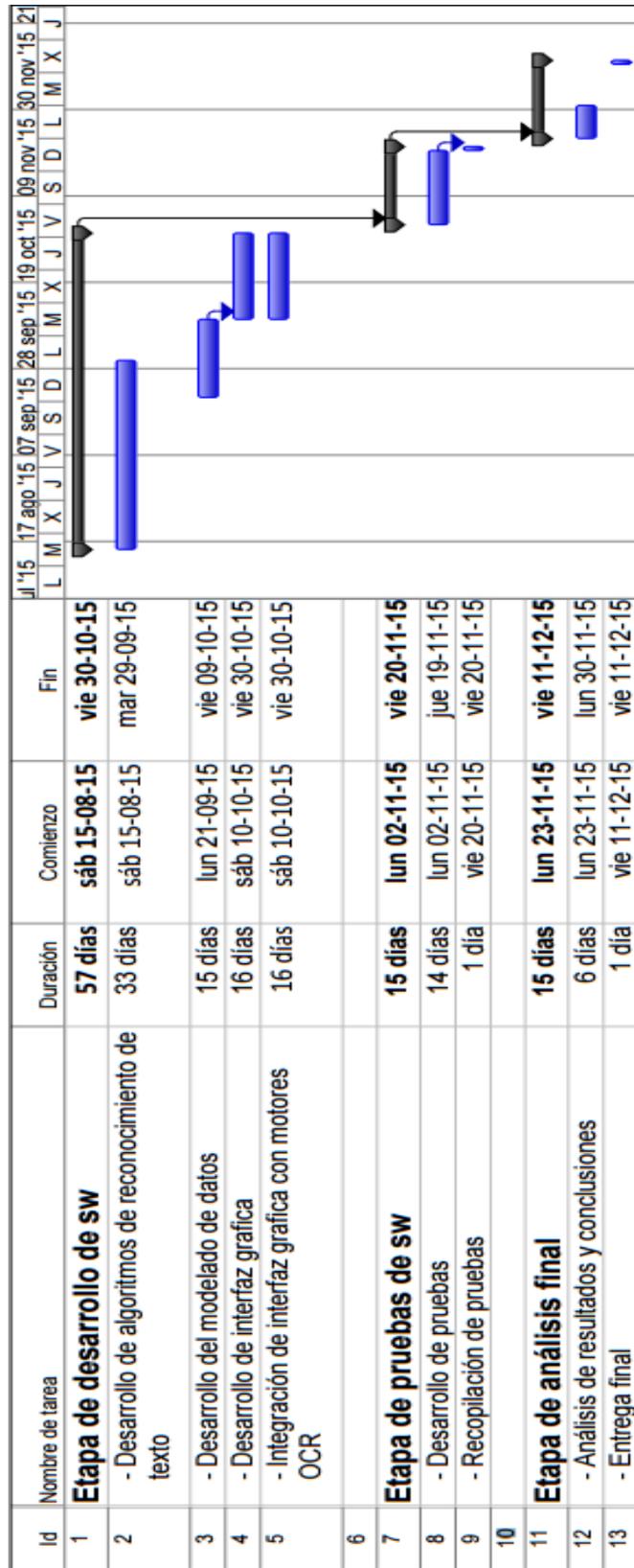


Figura 4-3 Carta Gantt etapa desarrollo

4.3 Detalle de las actividades

En las siguientes tablas se analiza en detalle cada actividad realizada en sus respectivas etapas.

Tabla 4-1 Etapa de planificación

Definir generalidad del proyecto: Se analiza la idea general del proyecto y su factibilidad.	3 día
Reunión con profesor guía: Se planifica una reunión para definir el tema y analizar pasos a seguir.	1 día
Entrevista con cliente: Se programa una entrevista para analizar problemáticas con un posible cliente	1 día
Construcción y revisión de carta Gantt: Se realiza un plan de trabajo para las distintas tareas a realizar	3 días
Redacción de informe de presentación: Se desarrolla el informe de presentación del proyecto	5 días
Evaluación de progreso: Se evalúa el progreso de la etapa de planificación y se analizan los pasos a seguir.	2 días

Tabla 4-2 Etapa de análisis de requerimientos

Diseñar formato de entrevista: Se diseña el formato de la entrevista para analizar los posibles problemas de la institución analizada	3 días
Reunión con IFIS: Se realiza la entrevista para analizar requerimientos, problemas y soluciones.	1 día
Definir requerimientos del sistema: Luego de la entrevista se definen los requerimientos del IFIS a partir de sus necesidades.	5 días
Segunda reunión: Esta reunión se define para hacer una reestructuración de los requerimientos y analizar funcionalidades	1 día
Definir funcionalidades: Luego de la segunda reunión se definen a cabalidad las funcionalidades del sistema.	3 días
Redacción de informe de avance: Se realiza la redacción del informe de avance donde se definen los alcances y propósitos del proyecto.	10 días
Reestructuración de carta Gantt: Se analizan las actividades del plan de trabajo por posibles modificaciones en las fechas establecidas	1 días
Evaluación de avance: Finalmente se evalúa el progreso de la etapa de análisis	1 día

Tabla 4-3 Etapa de análisis de sistema

Desarrollo de casos de uso: Se desarrollan los diagramas de casos de uso para mejorar la explicación de los módulos que manejan los usuarios.	4 días
Desarrollo de diagramas de secuencia: Se desarrollan los diagramas de secuencia para	5 días
Análisis de datos: Desarrollo del esquema de datos del sistema.	4 días
Modelo beta de interfaz: Se desarrolla una interfaz inicial del sistema para entender su representación.	5 días
Evaluación de análisis: Se evalúa el progreso de la etapa de diseño	1 días

Tabla 4-4 Etapa de diseño del sistema

Diseño de datos: Se realiza la transformación obtenida del análisis de datos para definir el modelo de datos a utilizar en las bases de datos.	3 días
Diseño arquitectónico: Se evalúan y se crean las estructuras requeridas para construir el sistema.	4 días
Diseño de interfaz: Se crea la interfaz del sw y la comunicación que tiene con los usuarios del sistema.	5 días
Diseño a nivel de componentes: Se convierte el diseño de datos donde se crea un nivel de abstracción cercano al código, con diagramas y dibujos.	4 días
Evaluación de diseño: Se analiza el desarrollo de los diseños realizados en esta etapa y se aprueban.	2 días

Tabla 4-5 Etapa final de la primera fase

Desarrollo de informe final: Se realiza la redacción del informe final para su posterior entrega.	10 días
Revisión y aprobación de informe: Se revisa el informe para sus posibles modificaciones, luego se aprueba para la entrega.	3 días

Tabla 4-6 Etapa de desarrollo de software

Desarrollo del modelado de datos: Se realiza la creación de la base de datos y todas sus tablas.	15 días
Desarrollo de algoritmos de reconocimiento de texto: Se realizan algoritmos para reconocer documentos digitalizados.	33 días
Desarrollo de interfaz gráfica: Se realiza el desarrollo de la capa de presentación del sistema.	16 días
Integración de sistema con motores OCR: Se conecta el sistema de información con los motores de reconocimientos de texto.	16 días

Tabla 4-7 Etapa de pruebas de software

Desarrollo de las pruebas de software: Se realizan distintos tipos de pruebas para analizar los resultados que entregan las distintas funciones que se ejecutan en el sistema.	20 días
Recopilación de pruebas: Se recopilan todas las pruebas realizadas y se ordenan para su posterior análisis.	1 día

Tabla 4-8 Etapa de análisis de resultados y conclusiones

Análisis de resultados: Se analizan los resultados generales obtenidos en del proyecto y se sacan las conclusiones necesarias para un trabajo futuro de mejoramiento y posible impacto al mercado.	7 días
Entrega Final: Se hace la entrega final del proyecto con el análisis de todos los resultados, conclusiones sacadas y trabajos futuros.	1 día

5 Estudio de factibilidad

Es necesario realizar un estudio de factibilidad para determinar la infraestructura tecnológica, la capacidad técnica y hacer un análisis financiero, económico y social que implica en la implementación del sistema descrito anteriormente identificando los costos y beneficios del proyecto. Este análisis permitirá determinar las posibilidades de generar el sistema propuesto tomando en cuenta aspectos esenciales. El estudio de factibilidad fue clasificado en el estudio de cuatro áreas las cuales se describen a continuación:

5.1 Factibilidad técnica

Con este estudio de factibilidad técnica se busca evaluar si el software si puede desarrollarse y si tiene las capacidades técnicas requeridas para cada alternativa del diseño considerado. Analizando las técnicas de reconocimiento de imagen con motores OCR. De acuerdo a las tecnologías necesarias para la implementación de este sistema se evaluará en dos áreas: hardware y software. Los equipos deben contar como mínimo con los siguientes requerimientos:

Tabla 5-1 Descripciones de hardware

Hardware	
Notebook	Procesador Intel Celeron 1.5 Ghz Memoria RAM 2 GB Disco duro 300 GB Tarjeta de red
Escáner	Procesador 667 Mhz Memoria RAM 1 GB Conexión Ethernet

Tabla 5-2 Descripción de software

Software	
Sistema operativo	Windows 8

5.1.1 Motores OCR

El acrónimo OCR hace referencia al reconocimiento óptico de caracteres. Es la tecnología que permite convertir distintos tipos de documentos en un archivo digital con diferentes funcionalidades y manejo de sus datos. Estos archivos pueden ser documentos en papel escaneado, archivos en formato PDF o una simple imagen capturada con una cámara digital. Este proceso está generalmente dirigido a la digitalización de textos el cual se encarga de identificar automáticamente los caracteres de una imagen a través de símbolos que pertenecen a un determinado alfabeto, luego se almacena dicho texto en forma de datos. El OCR es un proceso complejo que exige gran concentración y destreza para lograr

alcanzar un nivel óptimo de rendimiento y calidad, estas herramientas han logrado automatizar todo tipo de digitalización manual en muchas instituciones.

5.1.2 Esquema básico de un algoritmo de OCR

El principal propósito de los algoritmos es poder diferenciar un texto de una imagen. Para lograrlo hay que efectuar cuatro etapas que se describen a continuación:

5.1.2.1 Binarización

La mayoría de los OCR tienen como base trabajar con una imagen binaria (dos colores) por lo tanto conviene transformar una imagen de escala de grises o una de color por una imagen en blanco y negro conservando las propiedades esenciales de la imagen. Para la transformación se escoge un umbral adecuado donde los píxeles que no lo superen se convertirán en negro y el resto en blanco. El resultado de este proceso es una imagen en blanco y negro donde quedan claramente marcados los contornos de los caracteres y símbolos de la imagen.

5.1.2.2 Fragmentación de imagen

Este es el proceso más costoso y necesario para el reconocimiento de caracteres. La fragmentación de una imagen permite la descomposición de un texto en diferentes entidades lógicas que deben ser suficientemente significativas para su reconocimiento. Una de las técnicas más simples para imágenes de niveles de grises consiste en determinar agrupamientos (“clusters”) a partir del histograma, de tal forma que se permita una clasificación de los píxeles en regiones homogéneas.

5.1.2.3 Adelgazamiento de componentes

Luego de aislar los componentes de la imagen se tiene que aplicar un proceso de adelgazamiento para cada uno de ellos, este procedimiento consiste en ir borrando los puntos de los contornos de cada componente para que se conserve su tipología, la eliminación de los puntos consigue que la imagen continúe teniendo las mismas proporciones que la original y así conseguir q no se deforme.

5.1.3 Comparación con patrones

En esta etapa se comparan los patrones obtenidos con los que están almacenados en la base de datos. Que un OCR funciona de buena manera depende en gran medida de esta etapa, para esto existen distintos métodos para llevar a cabo la comparación. Uno de los métodos más usados es el de proyección, en el cual se obtienen proyecciones verticales y horizontales del carácter, luego se comparan con el alfabeto de caracteres posibles hasta encontrar la máxima coincidencia. Existen otros métodos: métodos geométricos o estadísticos, métodos estructurales, métodos neuro-miméticos, métodos markovianos o métodos de Zadeh.

5.1.4 Servicios OCR

El reconocimiento óptico de caracteres se ha aplicado a muchos servicios que han introducido estos procesos para aumentar su rendimiento. A continuación se muestran algunas de las aplicaciones más destacables que usan OCR.

5.1.4.1 Reconocimiento de texto manuscrito

Esta herramienta todavía sigue siendo un desafío, ya que se presentan muchas dificultades a la hora de reconocer un texto manuscrito y la mayoría de los algoritmos no consiguen buenos resultados. Aun así existen técnicas que mejoran la exactitud de la interpretación, se puede llegar a comprender una frase cuando la hemos terminado de leer, esto implica una operación de niveles morfológicos, léxico y sintáctico que se consigue mediante robustos algoritmos de reconocimiento del habla continua.

5.1.4.2 Reconocimiento de patentes

A través de aplicaciones de radar es posible localizar la matrícula de un vehículo con condiciones de iluminación, perspectiva y entorno variables. En la etapa de fragmentación se buscan texturas similares a la de una patente y se aísla el área rectangular que forma la patente. Finalmente se aplica un proceso de clasificación sobre el conjunto de píxeles que pertenece a la patente.

5.1.4.3 Reconocimiento de datos estructurados

Se usan para digitalizar de forma masiva grandes cantidades de documentos estructurados catalogando automáticamente los documentos con los metadatos obtenidos y archivándolos en formato digital de manera indexada para facilitar su posterior búsqueda. En esta técnica la única desventaja que tiene es que se deben diseñar previamente las plantillas para tener una herramienta más efectiva.

5.1.5 Combinación de escaneado y OCR

La mayoría de los programas de OCR pueden escanear un documento y efectuar de inmediato el reconocimiento óptico, la única condición es que el programa esté conectado directamente con la computadora que lo ejecuta. Escanear y efectuar el OCR página a página es un método razonable cuando se trata de pocos documentos, pero resulta un trabajo muy largo para trabajos de gran cantidad de archivos.

Para trabajar con gran volumen de documentos es más rápido y eficaz escanear en primer lugar todos los documentos y luego aplicar el proceso de OCR a todas las páginas de una sola vez.

5.2 Factibilidad económica

El estudio de factibilidad económica incluye un análisis de costos y beneficios asociados con cada alternativa del proyecto. Como la universidad y el desarrollador cuentan con los equipos y recursos técnicos necesarios para construir el sistema, no fue requerido ningún tipo de inversión.

Aunque no exista inversión en los recursos que se necesitan para desarrollar el sistema, igual se requiere realizar los cálculos económicos que conlleva la realización del software, para ello, se efectúa una simulación del costo total del proyecto, que en definitiva, es la suma total del costo en hardware, software y costo de las personas que trabajan en el proyecto. Como el proyecto se elabora como un trabajo de título, el personal a cargo del desarrollo del software no implicó ningún gasto.

Tabla 5-3 Detalle de insumos

Producto	Cantidad	Valor
Notebook	1	\$200.000
Escáner	1	\$300.000

5.3 Factibilidad operacional

El estudio de factibilidad operativa nos permite comprender la probabilidad de que se usa de buena forma un nuevo sistema, para esto se deben considerar tres aspectos fundamentales:

- 1 Un nuevo sistema puede ser muy complejo de utilizar para los usuarios del instituto de física, si esto ocurre los usuarios tienden a ignorar el sistema o bien usarlo de tal forma que cuse errores o fallas en el mismo.
- 2 Un nuevo sistema puede hacer que los usuarios se resistan a él por miedo a ser desplazado, intereses en el sistema antiguo, u otras razones.
- 3 Un nuevo sistema puede llegar a introducir cambios demasiados rápidos que no dejan adaptarse al personal adaptarse a él.

Desde el punto de vista operacional, se cree que el impacto del nuevo sistema, será positivo y sin muchas complicaciones las cuales se detallarán a continuación:

- Tiene un carácter corporativo, integrando a los usuarios con distintas funcionalidades.
- Es un sistema claro y sencillo de utilizar para usuarios inexpertos en el reconocimiento de texto.
- El sistema por ningún motivo es una amenaza para reemplazar a un trabajador, ya que el programa cumple una función de apoyo en el trabajo diario del personal del instituto de física.

5.4 Factibilidad legal

Para el desarrollo del software se utilizarán solo tecnologías que no infrinjan las leyes de derecho de autor o algún delito informático que viole condiciones de licencia. Además como este proyecto se realiza en las dependencias de la Escuela de Ingeniería Informática de la Pontificia Universidad Católica de Valparaíso donde los equipos cuentan con las licencias respectivas.

Para considerar la factibilidad legal se consideró la legislación actual respecto a delitos informáticos en el país, la Ley N° 19223 plantea cuatro artículos a tener en consideración:

- Artículo 1º

El que maliciosamente destruya o inutilice un sistema de tratamiento de información o sus partes o componentes, o impida, obstaculice o modifique su funcionamiento, sufrirá la pena de presidio menor en su grado medio a máximo. Si como consecuencia de estas conductas se afectaren los datos contenidos en el sistema, se aplicará la pena señalada en el inciso anterior, en su grado máximo.

- Artículo 2º

El que con el ánimo de apoderarse, usar o conocer indebidamente de la información contenida en un sistema de tratamiento de la misma, lo intercepte, interfiera o acceda a él, será castigado con presidio menor en su grado mínimo a medio.

- Artículo 3º

El que maliciosamente altere, dañe o destruya los datos contenidos en un sistema de tratamiento de información, será castigado con presidio menor en su grado medio.

- Artículo 4º

El que maliciosamente revele o difunda los datos contenidos en un sistema de información, sufrirá la pena de presidio menor en su grado medio. Si quien incurre en estas conductas es el responsable del sistema de información, la pena se aumentará en un grado.

6 Análisis y gestión de riesgo

El análisis y gestión del riesgo nos permite hacer un estudio de posibles consecuencias no deseadas con impactos negativos en nuestro sistema. Los riesgos del proyecto son una amenaza para el plan de trabajo diseñado, provocando retrasos y que los costos aumenten. Es por esto que se hará la identificación de los posibles riesgos con sus respectivas evaluaciones de probabilidad e impacto, pudiendo evitar posibles problemas sabiendo mitigar los riesgos que se describen a continuación.

6.1 Riesgo de la tecnología a construir

Debido a que se estudiarán nuevas herramientas como son los motores de reconocimiento de caracteres existe el riesgo de que se dificulte el rápido aprendizaje de las nuevas técnicas de programación, lo que tendría un impacto en los tiempos del proyecto. Para evitar este riesgo se hará un estudio enfocado en cómo trabajan las tecnologías de los motores OCR, distintos algoritmos de captura de información y librerías que utilicen técnicas de reconocimiento de textos.

6.2 Riesgo de las características del cliente

Existe un eventual riesgo que ocurre en la mayoría de los proyectos informáticos. El riesgo asociado con la comunicación y el entendimiento que desarrollamos con el cliente, si existe una mala comunicación se producen deficiencias en la explicación de los requerimientos del sistema, por otra parte puede haber una buena comunicación pero no un buen entendimiento de lo requerido, lo que provocaría un altísimo impacto en la calidad final del software. Para contrarrestar este riesgo se trabaja para lograr una buena comunicación, coordinando continuas reuniones, utilizando diagramas y modelando situaciones para la buena explicación y el buen entendimiento de las partes.

6.3 Riesgo asociado a la continuidad del software

Existe el riesgo que el software no se adapte a cambios radicales en los formatos de los documentos, ya que se crean algoritmos específicos para cierto tipo de documentos. Tendría un gran impacto ya que se tendría que volver a programar para hacer un nuevo tipo de captura de la información. Para prevenir este riesgo se pedirá expresamente analizar los documentos correspondientes a los años más recientes y pedir que se mantenga el formato para evitar problemas en el reconocimiento.

7 Análisis de la solución

7.1 Arquitectura del sistema

Se decide utilizar una arquitectura lógica de tres capas. Para explicarlo de mejor manera se muestra el siguiente diagrama:

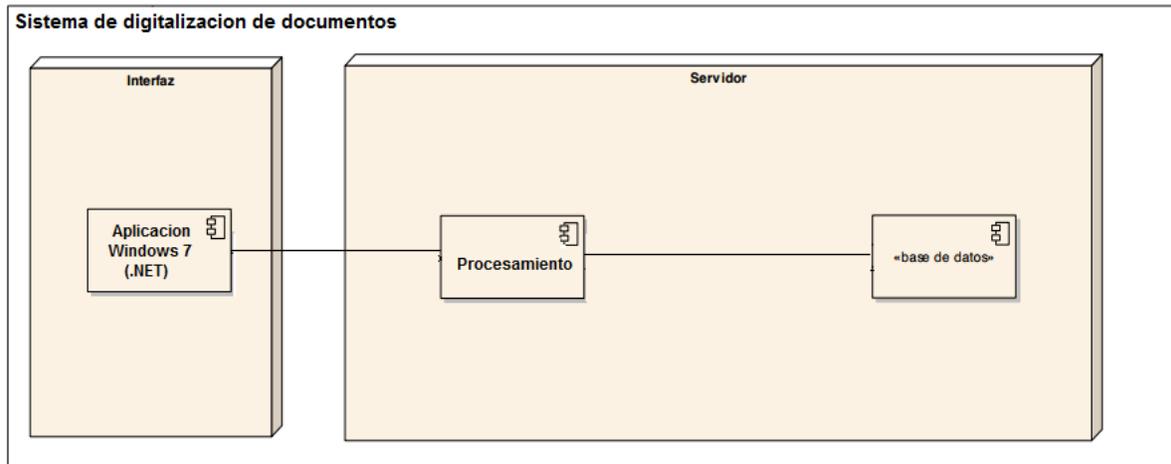


Figura 7-1 Arquitectura del sistema

Podemos ver la estructura de las 3 capas:

- La capa de presentación corresponde a la interfaz principal del sistema y es donde el usuario hace toda la interacción con el programa.
- Dentro del servidor se encuentra la capa de negocio donde se reciben las peticiones del usuario y se hace el procesamiento de los datos utilizando herramientas de reconocimiento de caracteres.
- En la capa de datos es donde se almacenan todos los datos del sistema utilizando bases de datos avanzada.

7.2 Diagrama relacional

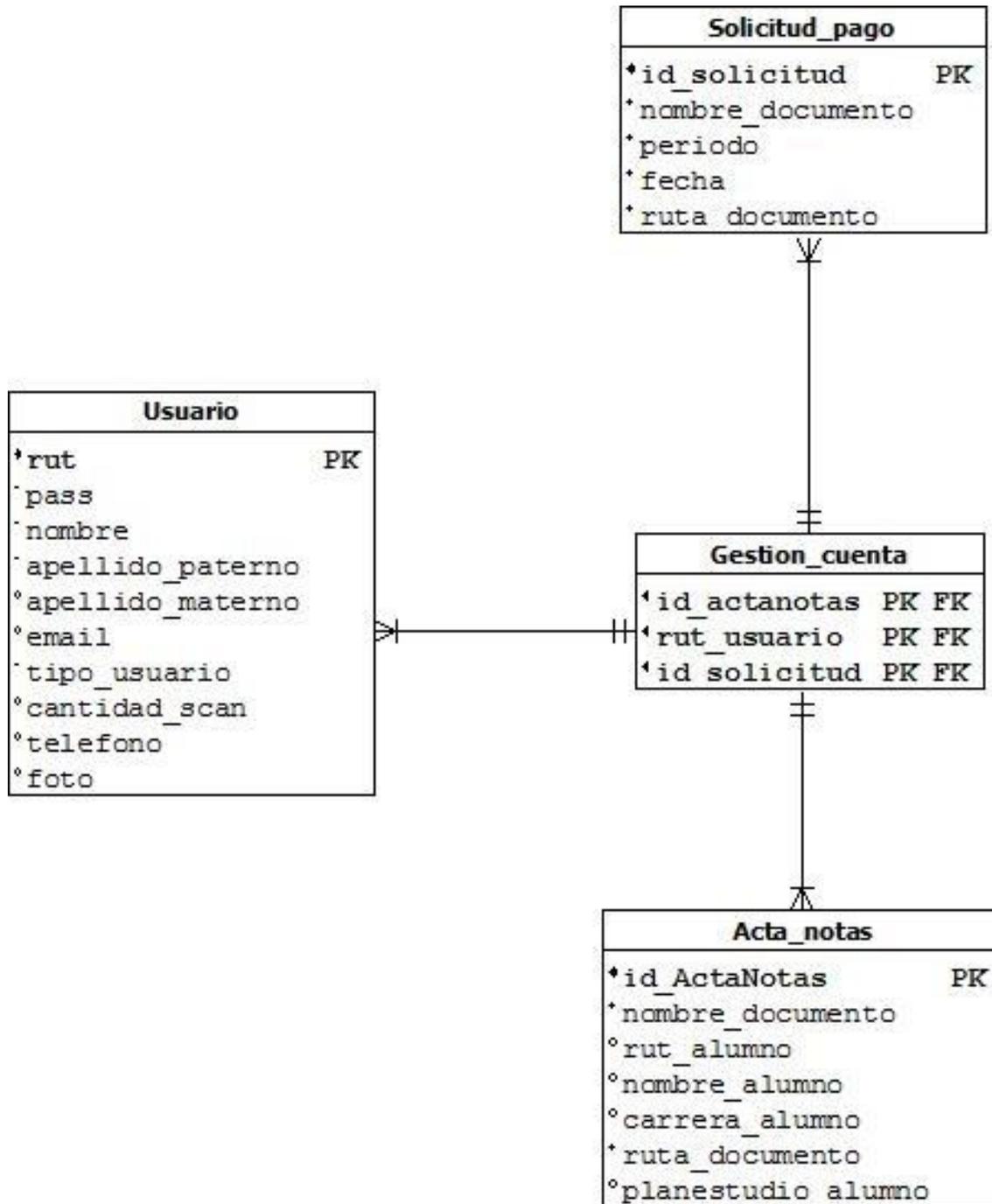


Figura 7-2 Diagrama Relacional

7.3 Diagrama de clases

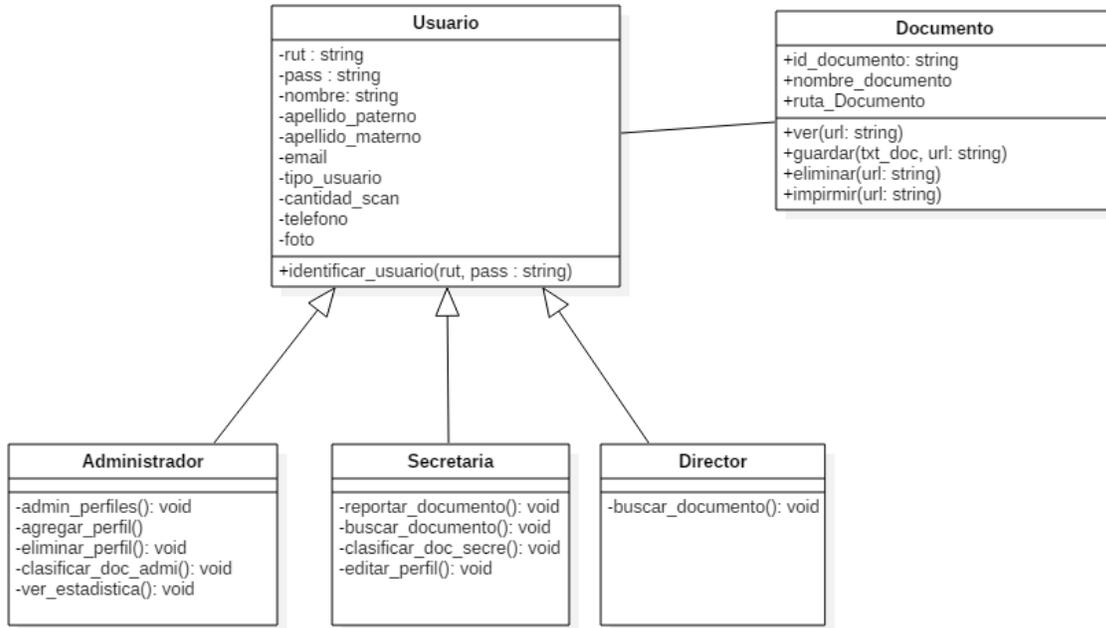


Figura 7-3 Diagrama de clases

7.4 Caso de uso general del sistema

En el siguiente diagrama se muestra el resultado de reuniones con el IFIS donde explicaron en detalle sus requerimientos para que las funcionalidades del sistemas cumplan con todas las exigencias que se pretenden desarrollar a los largo del proyecto.

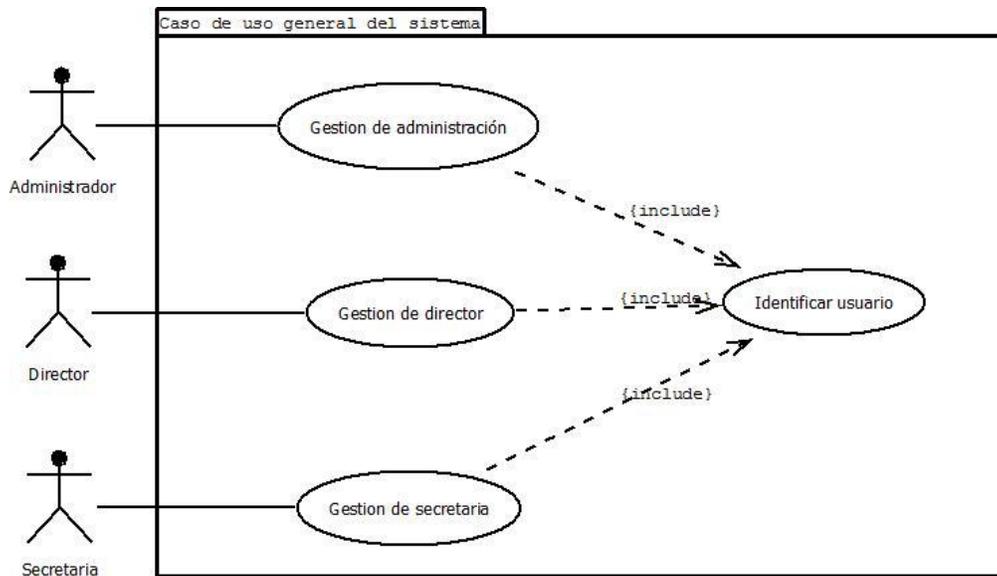


Figura 7-4 Diagrama de caso de uso general del sistema

A continuación se presentan los siguientes diagramas para facilitar la explicación de cómo interactúa cada actor del sistema con sus respectivos módulos.

7.5 Caso de uso administrador

Cuando el usuario Administrador inicia su sesión ingresa en un módulo donde tiene la opción de administrar las cuentas de usuarios registradas en el sistema, las cuales tiene funcionalidades como ver, agregar, modificar y eliminar los perfiles. Otra funcionalidad importante del administrador es revisar los reportes y estadísticas. Por una parte las estadísticas del sistema permiten visualizar que usuario utiliza con mayor frecuencia el sistema y la cantidad de cada documento que fue escaneado en un periodo de tiempo. Y por otra parte la funcionalidad de ver los archivos reportados permite analizar los documentos que por algún motivo el motor de reconocimiento no tuvo el resultado esperado, por lo que el administrador podrá revisar e ingresar los datos correctos al sistema.

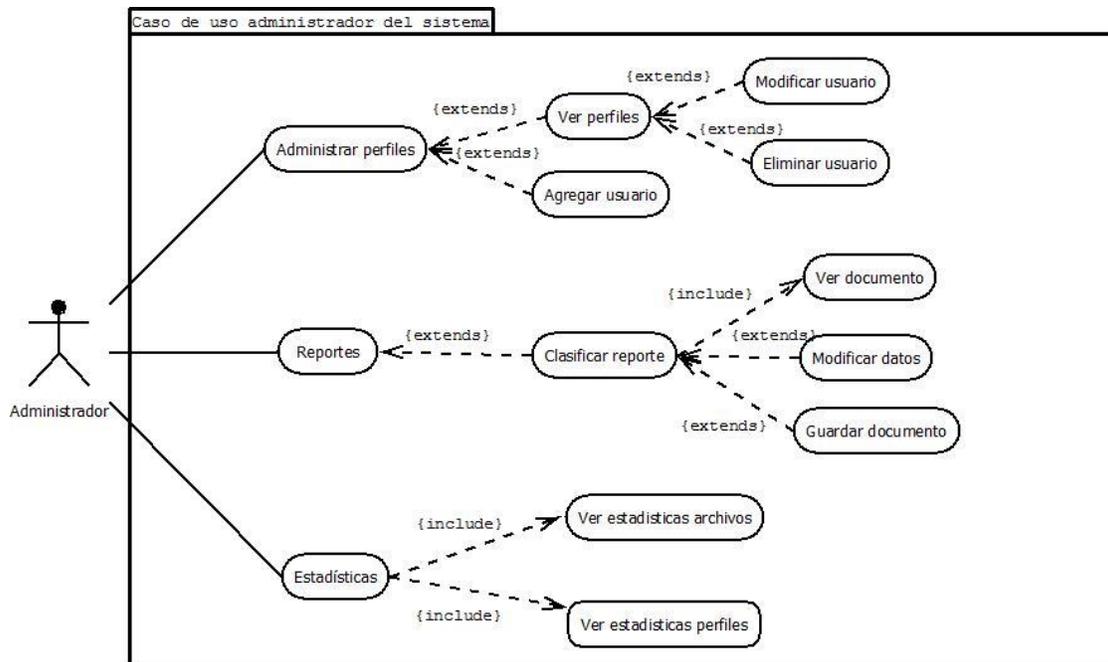


Figura 7-5 Diagrama caso de uso módulo administrador

7.6 Caso de uso director

Para el módulo correspondiente al usuario director podemos contemplar opciones simples que son solo para visualizar los documentos digitalizados, por un lado tenemos la opción de buscar de manera avanzada un archivo, buscando por nombre, año, etc. Y otra opción de analizar las carpetas existentes. Al momento de encontrar el documento que se necesita tenemos tres nuevas alternativas para hacer con el archivo, estas son mostrar documentos, imprimir documentos, guardar documentos.

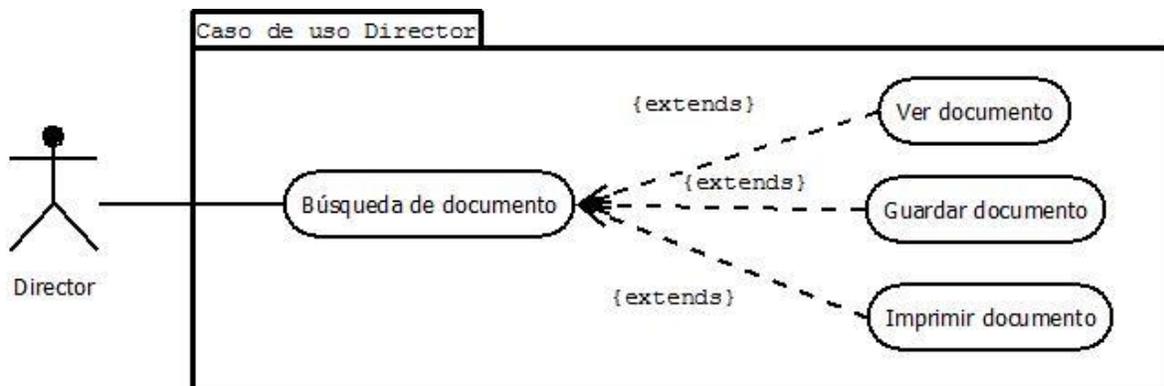


Figura 7-6 Diagrama caso de uso módulo director

7.7 Caso de uso secretaria

El módulo de la secretaria es uno de los más importantes y es el encargado de satisfacer la necesidad por la que fue diseñado el sistema. La secretaria es el usuario más

activo del sistema cumpliendo la función de analizar los archivos escaneados, teniendo la posibilidad de clasificarlos de acuerdo al reconocimiento de datos o reportarlo si fue reconocido de mala manera. También tiene otras opciones como buscar documentos de manera avanzada o simple, al momento de encontrar el documento tiene opciones de imprimir, ver, almacenar o eliminar. Además cuenta con funciones en su propio perfil de usuario, puede modificar sus datos personales.

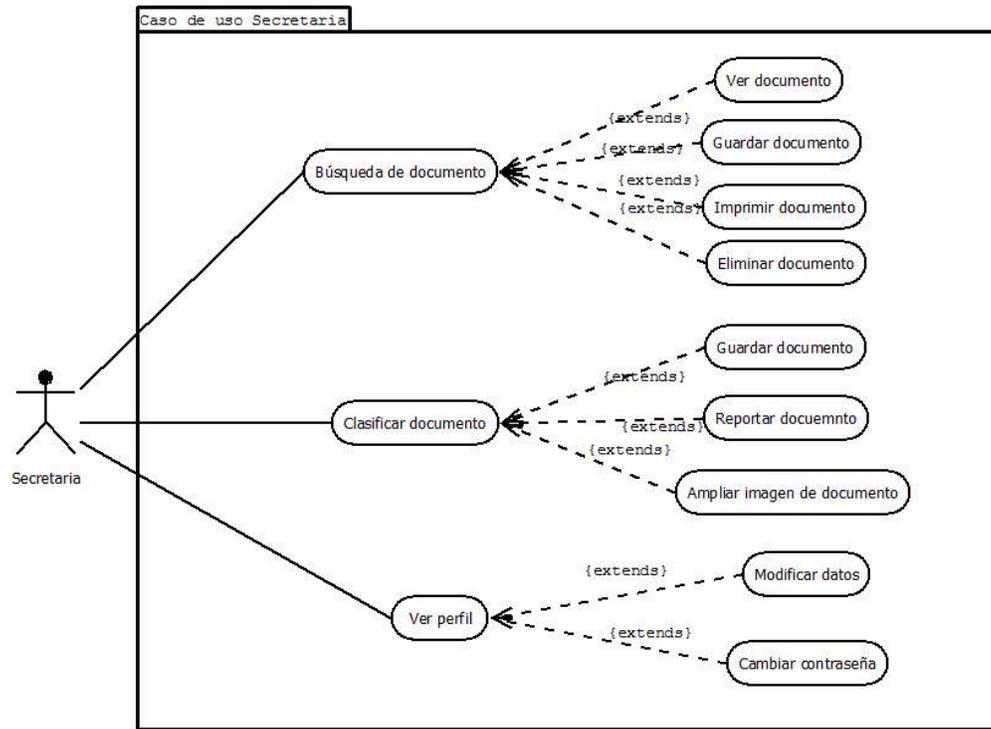


Figura 7-7 Diagrama caso de uso módulo secretaria

7.8 Caso de uso narrativo del sistema inicial

Tabla 7-1 Caso de uso narrativo del sistema inicial

Caso de uso	Sistema general
Actor Principal	Usuarios (secretaria, administrador, director)
Participantes e Intereses	Sistema y usuario: tiene como finalidad identificar al usuario e ingresar a su respectivo módulo.
Precondiciones	Validación de datos para iniciar sesión
Poscondiciones	Ingreso de usuario correcto con sus respectivas funcionalidades.
Escenario Principal	<ol style="list-style-type: none"> 1.- El usuario ingresa a la pantalla inicial donde se ingresa los datos rut y contraseña de usuario para el inicio de sesión. 2.- El Sistema verifica si los datos son correctos 3.- El sistema confirma los datos y hace el ingreso al módulo correspondiente al usuario identificado.
Extensiones	3.1.- Usuario inválido envía mensaje con usuario incorrecto.
Requisitos Especiales	Debe ingresar un rut válido
Frecuencia de Ocurrencia	Muy Alta

7.10 Caso de uso narrativo Módulo Secretaria

Tabla 7-2 Caso de uso narrativo módulo secretaria

Caso de uso	Módulo Secretaria
Actor Principal	Secretaria
Participantes e Intereses	Gestionar el proceso del manejo de archivos
Precondiciones	Usuario previamente autenticado
Poscondiciones	Usuario con funcionalidades de manejo de archivos
Escenario Principal	1.- El usuario ingresa a la pantalla inicial viendo todas sus funcionalidades. 2.- Búsqueda de documentos 3.- Clasificar documento 4.- Vista de perfil
Extensiones	2.1.- El usuario podrá hacer una búsqueda rápida o avanzada de un documento en específico. 4.1.- Vista de perfil con opciones de modificar datos y contraseña
Requisitos Especiales	No se pueden modificar el rut del usuario, ni el tipo de usuario
Frecuencia de Ocurrencia	Muy Alta

7.11 Caso de uso narrativo Módulo Administrador

Tabla 7-3 Caso de uso narrativo módulo administrador

Caso de uso	Modulo Administrador
Actor Principal	Administrador del sistema
Participantes e Intereses	El administrador gestiona el funcionamiento general del sistema y a sus usuarios.
Precondiciones	Validación de datos para el inicio de sesión
Poscondiciones	Ingreso de usuario correcto con sus respectivas funcionalidades.
Escenario Principal	<ol style="list-style-type: none"> 1.- El usuario ingresa a la pantalla inicial donde se le muestran las funcionalidades del sistema. 2.- El usuario podrá administrar los perfiles del sistema. 3.- Ver detalles de estadísticas del sistema. 4.- Ver reportes de documentos con la opción de modificarlos y clasificarlos.
Extensiones	<ol style="list-style-type: none"> 2.1.- Ver un perfil de usuario 2.2.- Modificar un perfil de usuario 2.3.- Agregar un nuevo perfil de usuario 2.4.- Eliminar un usuario 3.1 Ver reportes hecho por secretarias de documentos mal reconocidos.
Requisitos Especiales	-----
Frecuencia de Ocurrencia	Media

7.13 Caso de uso narrativo Módulo Director

Tabla 7-4 Caso de uso módulo director

Caso de uso	Modulo director
Actor Principal	Director
Participantes e Intereses	Director de sistema con funcionalidades de visualización de documentos digitales.
Precondiciones	Validación de datos para iniciar sesión
Poscondiciones	Ingreso de usuario correcto con sus respectivas funcionalidades.
Escenario Principal	1.- Búsqueda de documentos
Extensiones	1.1.- Ver documento 1.2.- Guardar documento 1.3- Imprimir documento
Requisitos Especiales	----- -
Frecuencia de Ocurrencia	Media

8 Diagramas de secuencia

Los diagramas de secuencia que se presentan a continuación tienen por finalidad modelar la interacción entre los objetos del sistema.

8.1 Diagrama de secuencia Ingresar Usuario

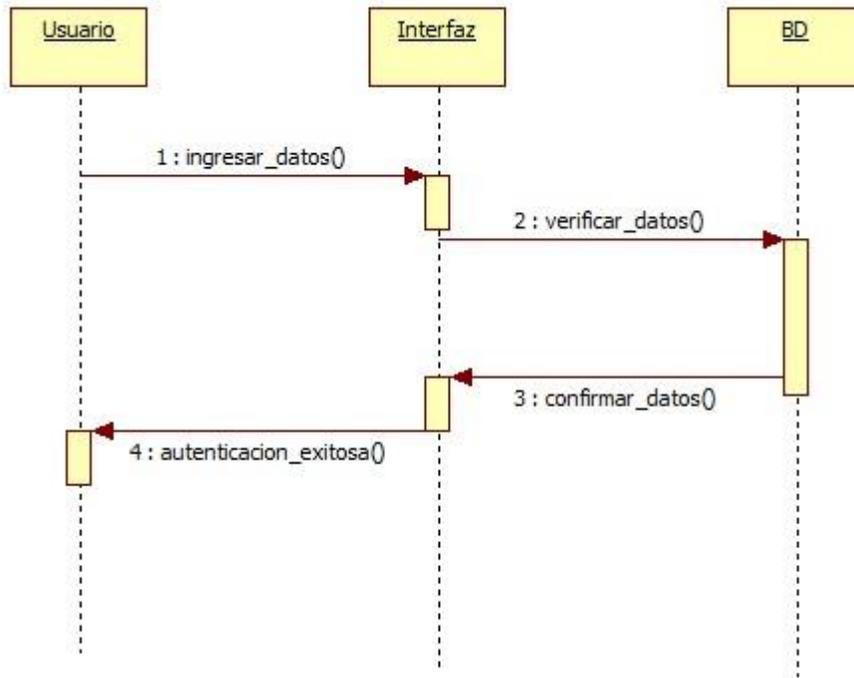


Figura 8-1 Diagrama de secuencia iniciar sesión

8.2 Diagrama de secuencia Clasificar/Reportar Documento

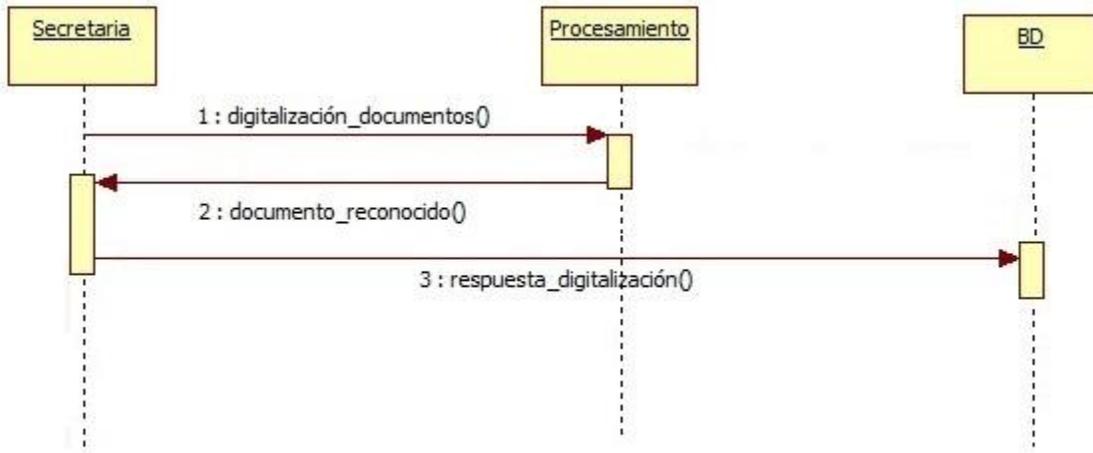


Figura 8-2 Diagrama de secuencia clasificar/reportar documento

8.3 Diagrama de secuencia Buscar Documento

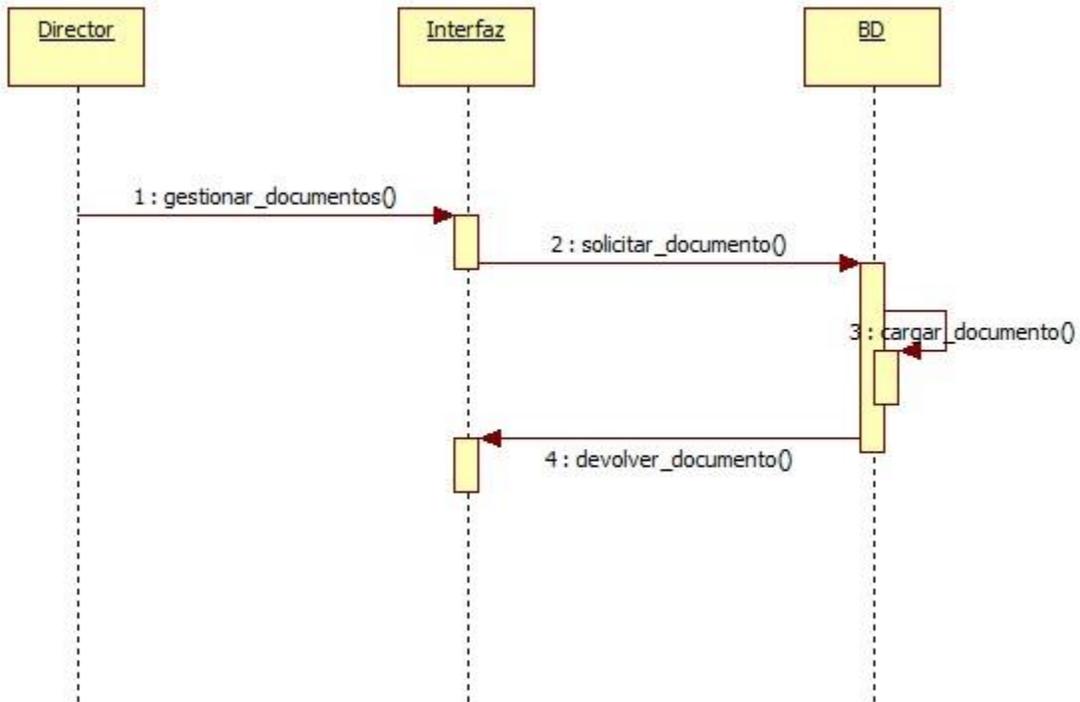


Figura 8-3 Diagrama de secuencia buscar documento

9 Prototipos gráficas

A continuación, se detallaran de manera gráfica las funcionalidades del sistema mostrando los prototipos finales del sistema. Se explicaran en detalle los módulos de los tres actores y su implementación en el sistema.

9.1 Interfaz de acceso al sistema

Es la pantalla inicial del sistema y verifica al usuario que desea ingresar consultando los datos esenciales que son rut y contraseña, luego de verificar al usuario ingresa a su respectivo módulo.



Figura 9-1 Acceso al sistema

9.2 Interfaz módulo director

Este es el módulo que menos interactúa en el sistema, manejando acceso a los archivos y sus datos con funciones de visualización, almacenamiento e impresión de documentos encontrados a través de búsquedas simples o avanzadas.

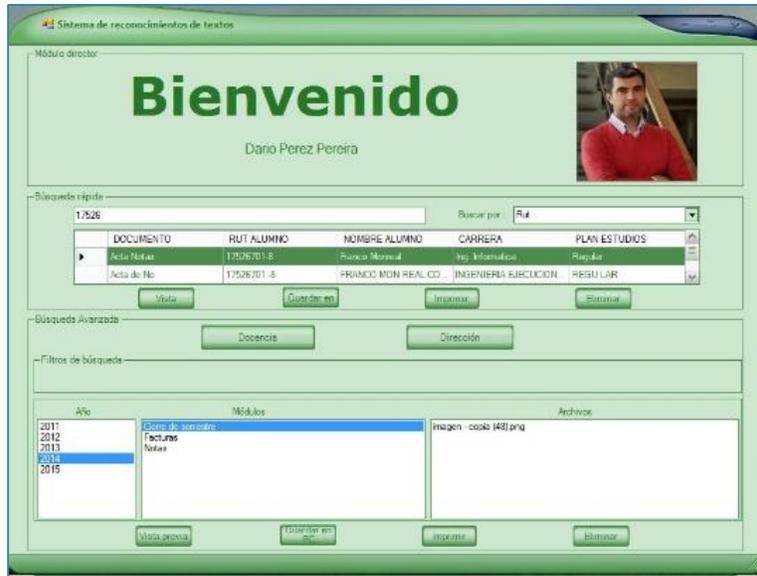


Figura 9-2 Interfaz director

9.3 Interfaz secretaria

Similar a la interfaz del director, añadiendo opciones de clasificar documentos, donde se define si un documento fue reconocido de buena forma, también puede editar su perfil y buscar documentos del sistema de manera simple y avanzada.

9.3.1 Clasificar documentos

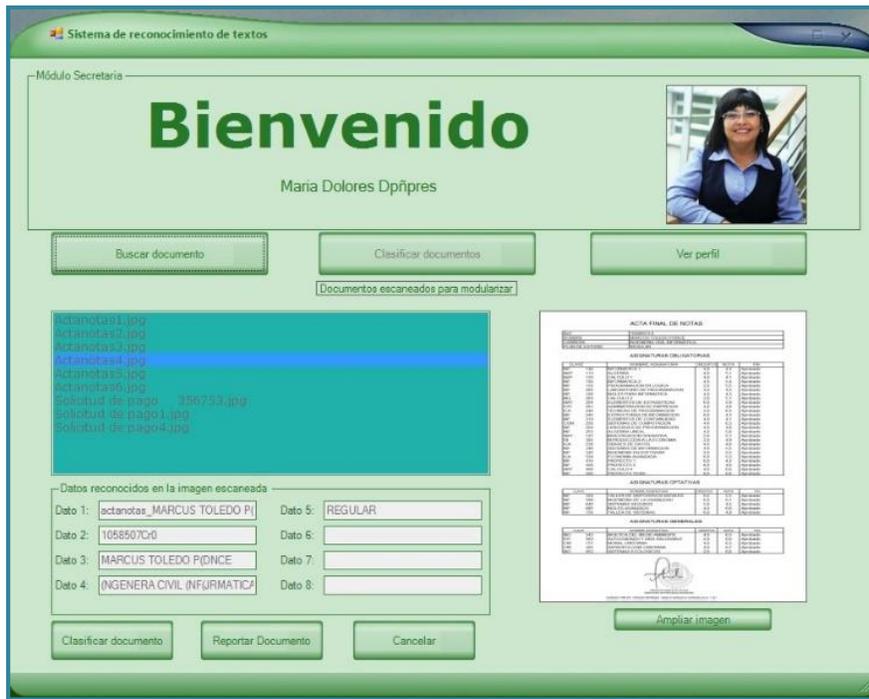


Figura 9-3 Interfaz clasificar documentos secretaria

9.3.2 Ver Perfil

Sistema de reconocimiento de textos

Módulo: Secretaria

Bienvenido

Maria Dolores Dpripres

Buscar documento Clasificar documentos Ver perfil

DATOS DE USUARIO

Datos personales

NOMBRES	Maria
APELLIDO PATERNO	Dolores
APELLIDO MATERNO	Dpripres
RUT	1234
EMAIL	@dsda
TIPO DE USUARIO	Secretaria
TELEFONO	123312123

MODIFICAR CAMBIAR CONTRASEÑA

Figura 9-4 Interfaz ver perfil

9.4 Interfaz administrador

El módulo de administrador se encarga de administrar las cuentas de los usuarios, con opciones de modificar, agregar o eliminar un usuario determinado. También maneja datos estadísticos del sistema y además es el encargado de analizar los reportes de documentos mal reconocidos.

9.4.1 Administrar perfiles



Figura 9-5 Interfaz administrar perfiles administrador

9.4.2 Ver perfil de otro usuario



Figura 9-6 Interfaz ver perfil de secretaria

9.4.3 Agregar Usuario

The screenshot shows a web interface for adding a user. At the top, it says 'Sistema de reconocimientos de textos' and 'Módulo Administrador del sistema'. A large green 'Bienvenido' message is displayed, along with the name 'Francisco Zamora Palacios' and a small profile picture. Below this, there is a section for 'Datos personales' with several input fields: 'NOMBRES', 'APELLIDO PATERNO', 'APELLIDO MATERNO', 'RUT', 'PASS', 'EMAIL', 'TIPO DE USUARIO' (a dropdown menu), and 'TELEFONO'. To the left of these fields is a placeholder for a profile picture with the text '(click para agregar nueva imagen)'. At the bottom, there are two buttons: 'Guardar' and 'Cancelar'.

Figura 9-7 Interfaz agregar usuario

9.4.4 Reportes de documentos

The screenshot shows a web interface for reporting documents. At the top, it says 'Sistema de reconocimiento de textos' and 'Módulo Administrador del sistema'. A large green 'Bienvenido' message is displayed, along with the name 'Francisco Zamora Palacios' and a small profile picture. Below this, there are three buttons: 'Administrar perfiles', 'Reportes', and 'Estadísticas'. A blue bar indicates the current document being viewed: 'Actanotas4.jpg'. Below this, there are two main sections: 'Acta de notas' and 'Solicitud de pago'. The 'Acta de notas' section has several input fields: 'Nombre documento' (with the value 'actanotas_MARCUS TOLEDO P(DNCE)'), 'Rut. Alumno' (with the value '1058507G0'), 'Nombre alumno' (with the value 'MARCUS TOLEDO P(DNCE)'), 'Carrera' (with the value 'INGENERIA CIVIL (NFJ/RMATICA)'), and 'Plan de estudio' (with the value 'REGULAR'). The 'Solicitud de pago' section has input fields for 'Numero de documento', 'Nombre documento', 'Periodo', and 'Fecha'. Both sections have a 'Guardar' button at the bottom.

Figura 9-8 Interfaz de documentos reportados

9.4.5 Estadísticas del sistema

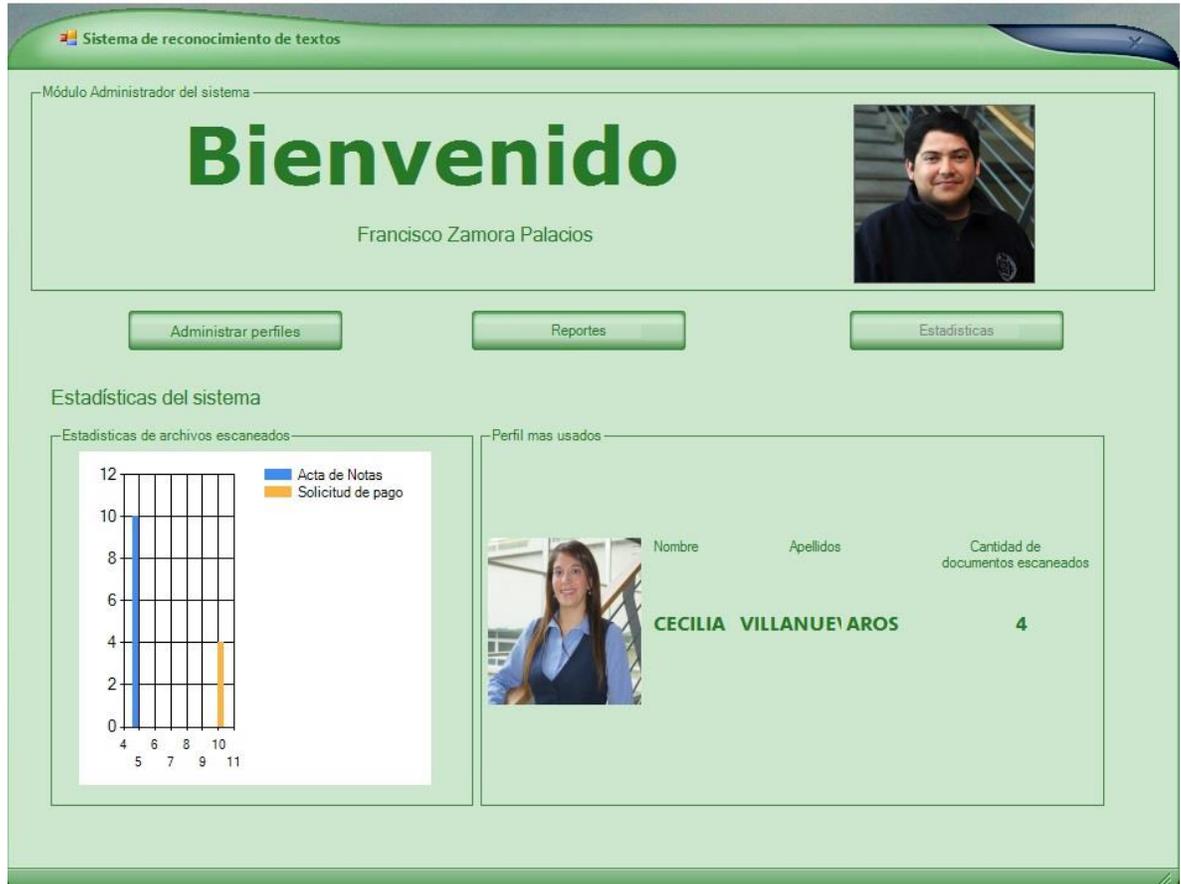


Figura 9-9 Interfaz estadísticas del sistema

10 Conclusión

En conclusión podemos destacar lo importante que es para una institución la digitalización de documentos, haciendo posible un respaldo de los archivos que por años se guardan en bodegas haciendo uso de espacio físico innecesario y con el riesgo constante de perder la información por cualquier circunstancia, ya sea catástrofe natural o alguna provocada por el humano.

Debemos tener en cuenta que con las nuevas tecnologías es posible almacenar de manera digital toda la información rescatada de un documento físico, permitiendo hacer búsquedas rápidas sin necesidad de ir leyendo archivo por archivo hasta encontrar la información. Con ayuda de las técnicas de reconocimiento de texto podemos hacer que el trabajo humano de digitalizar manualmente la información sea de manera automatizada con algoritmos complejos para identificar de buena manera el documento y asociarlo a los módulos correspondientes.

Finalmente el desarrollo del sistema obtuvo resultados positivos en todo aspecto, cuenta con una buena interfaz de usuario, la cual permite organizar a los distintos tipos de trabajadores que ingresaran a los módulos correspondientes. Además los resultados del reconocimiento de texto generalmente obtienen resultados con tasas de acierto de más del 75%. Esto se reduce en una satisfacción a la hora de usar el sistema por parte de los usuarios, los cuales confían en el reconocimiento de texto.

El sistema es una excelente herramienta para todo tipo de institución con las mismas problemáticas del Instituto de Física, ya que se podrá reducir notablemente el tedioso trabajo de la digitalización manual de los datos perdiendo tiempo valioso. Es por esto que el sistema a pesar de actualmente realizar la digitalización de dos documentos específicos, se debe analizar de una manera más global, ya que puede entregar soluciones informáticas a muchas instituciones que tengan problemas similares y que deseen contar con un sistema de automatización de procesos de capturas de información a través de motores OCR.

11 Referencias

[Espinoza, SIA] Fernando Espinoza F. Disponible vía web campuscurico.utalca.cl/~fespinos/20-sistemas_informacion_presentacion.pdf revisada por última vez el 25 de septiembre del 2014.

[OCR] Acerca de OCR. Disponible vía web <http://latam.abbyy.com/finereader/ocr/> revisada por última vez el 25 de septiembre del 2014.

[Ciampagna, SIG] José Ciampagna, Los sistemas de información en la organización. Disponible vía web elprofejose.files.wordpress.com/2011/10/03_los-sistemas-de-informacic3b3n-en-la-organizacion.pdf

[Dpto ciencia, SIA] Departamento de ciencias de la computación, SIA fundamentos de diseño de base de datos Disponible vía web www.kulturklik.euskadi.net/wp-content/uploads/2011/04/pildora-OCR-2.pdf

[Guerra, 14] Lisbet Guerra, Eriberto Venegas. Software libre para la restauración y transcripción de imágenes. Disponible vía web www.3ciencias.com/wp-content/uploads/2014/02/DOCLUX-OCR.pdf revisada por última vez el 19 de noviembre del 2014.

[Marín] OCR. NET Motor para el reconocimiento óptico de caracteres en .NET. Disponible vía web <http://sergiogonzalezc.files.wordpress.com/2011/04/dnm041.pdf> revisada por última vez 19 de noviembre del 2014.

[Pressman, 03] Roger Pressman. Ingeniería del software. Junio 2003.

12 Anexos

12.1 Acta final de notas

ACTA FINAL DE NOTAS

RUT	15709336-3
NOMBRE	CATALINA BELEN MORALES MOREIRA
CARRERA	INGENIERIA EJECUCION INFORMATICA
PLAN DE ESTUDIO	REGULAR

ASIGNATURAS OBLIGATORIAS

CLAVE	NOMBRE ASIGNATURA	CREDITOS	NOTA	FIN.
INF 140	INFORMATICA 1	4,0	4,2	Aprobado
MAT 113	ALGEBRA	4,0	5,1	Aprobado
MAT 123	CALCULO 1	4,0	4,1	Aprobado
INF 150	INFORMATICA 2	4,0	5,4	Aprobado
INF 152	PROGRAMACION EN LOGICA	2,0	5,5	Aprobado
INF 265	LABORATORIO DE PROGRAMACION	4,0	4,5	Aprobado
INF 100	INGLES PARA INFORMATICA	4,0	4,3	Aprobado
ING 203	CALCULO 2	2,0	5,1	Aprobado
MAT 204	ELEMENTOS DE ESTADISTICAS	6,0	5,9	Aprobado
EST 201	ADMINISTRACION DE EMPRESAS	4,0	4,8	Aprobado
ICA 240	TECNICAS DE PROGRAMACION	2,0	6,3	Aprobado
INF 245	ESTRUCTURAS DE INFORMACION	6,0	4,3	Aprobado
INF 214	ELEMENTOS DE CONTABILIDAD	4,0	4,1	Aprobado
COM 250	SISTEMAS DE COMPUTACION	4,0	6,3	Aprobado
INF 252	LENGUAJES DE PROGRAMACION	4,0	4,8	Aprobado
INF 253	ALGEBRA LINEAL	4,0	5,8	Aprobado
MAT 127	INVESTIGACION OPERATIVA	5,0	5,1	Aprobado
EII 305	INTRODUCCION A LA ECONOMIA	2,0	4,9	Aprobado
ICA 230	SBASES DE DATOS	4,0	4,6	Aprobado
INF 340	SISTEMAS DE INFORMACION	4,0	5,5	Aprobado
INF 342	INGENIERIA EN SOFTWARE	2,0	5,5	Aprobado
ICA 524	ECONOMIA AVANZADA	6,0	5,3	Aprobado
INF 410	PROYECTO 1	6,0	4,2	Aprobado
INF 450	PROYECTO 2	8,0	4,6	Aprobado
MAT 400	CALCULO 4	4,0	6,6	Aprobado
INF 500	PROYECTO TESIS	8,0	6,8	Aprobado

ASIGNATURAS OPTATIVAS

CLAVE	NOMBRE ASIGNATURA	CREDITOS	NOTA	FIN.
INF 524	TALLER DE DISPOSITIVOS MOVILES	6,0	5,5	Aprobado
INF 599	INGENIERIA DE LA USABILIDAD	6,0	6,1	Aprobado
INF 642	SISTEMAS SEGUROS	5,0	4,5	Aprobado
INF 685	INGLES AVANZADO	4,0	6,6	Aprobado
INF 750	TALLER DE SISTEMAS	6,0	4,9	Aprobado

ASIGNATURAS GENERALES

CLAVE	NOMBRE ASIGNATURA	CREDITOS	NOTA	FIN.
BIO 543	BIOETICA DEL MEDIO AMBIENTE	4,0	6,5	Aprobado
EFI 563	AUTOCUIDADO Y VIDA SALUDABLE	2,0	6,6	Aprobado
CRI 157	MORAL CRISTIANA	4,0	6,3	Aprobado
CRI 325	ANTROPOLOGIA CRISTIANA	4,0	6,7	Aprobado
BIO 953	SISTEMAS ECOLOGICOS	2,0	6,8	Aprobado



ANGELICA CASALETTI LOYOLA
DIRECTORA DE PROCESOS DOCENTES

Certificado n°480.549 , Verificador 5877653a03 . Validar el certificado en certificados.ucv.cl - 1 de 1

Figura 12-1 Acta final de notas digitalizada

12.2 Solicitud de pago de factura

SOLICITUD DE PAGO DE FACTURAS N°
PERIODO **2014**

356753

ESTADO SOLICITUD	
Estado Formulario :	Pendiente

SOLICITANTE			
Solicitante:	Secretaria Instituto de Fisica	Fecha:	18 10 2014

DETALLE SOLICITUD	
Email Solicitante	darioifis@ucv.cl
Anexo Solicitante	4853
Cuenta Cargo	1432323
Total Cargo	1437176
Glosa / Descripcion	Equipamiento academico

DETALLE FACTURA					
RUT	NOMBRE	N° FAC	FECHA	DETALLE SERVICIO	TOTAL
654321-2	Electrolis CIA	123452	03-11-2014	Insumo Electricos	12341
654321-2	Electrolis CIA	234111	21 -11- 2014	Insumo Electricos	23321
654321-2	Electrolis CIA	432512	18-11-2014	Insumo Electricos	31231
TOTAL					66893
<p>NOTA:</p> <ol style="list-style-type: none"> 1. En caso de tramitar a pago una FACTURA MANUAL esta deberá ser remitida físicamente a la Direccion de finanzas junto a la copia/comprobante obtenida al autorizar/enviar el formulario respectivo a traves del Sistema Ventanilla Unica 2. Si dentro de un plazo de 2 dias habiles la documentacion fisica respectiva no es recibida a conformidad en la Direccion de Finanzas la solicitud realizada a traves del sistema será rechazada perdiendo vigencia la fecha de pago originalmente propuesta para la solicitud. 					

Figura 12-2 Solicitud de comprobante de pago digitalizada